

## King's Research Portal

DOI:

[10.1002/art.39560](https://doi.org/10.1002/art.39560)

Document Version

Peer reviewed version

[Link to publication record in King's Research Portal](#)

*Citation for published version (APA):*

Liu, K., Kurien, B. T., Zimmerman, S. L., Kaufman, K. M., Taft, D. H., Kottyan, L. C., Lazaro, S., Weaver, C. A., Ice, J. A., Adler, A. J., Chodosh, J., Radfar, L., Rasmussen, A., Stone, D. U., Lewis, D. M., Li, S., Koelsch, K. A., Igoe, A., Talsania, M., ... Hal Scofield, R. (2016). X Chromosome Dose and Sex Bias in Autoimmune Diseases: Increased Prevalence of 47,XXX in Systemic Lupus Erythematosus and Sjögren's Syndrome. *Arthritis and Rheumatology*, 68(5), 1290-1300. <https://doi.org/10.1002/art.39560>

### Citing this paper

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

### General rights

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

### Take down policy

If you believe that this document breaches copyright please contact [librarypure@kcl.ac.uk](mailto:librarypure@kcl.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.

**X Chromosome Dose and Sex Bias in Autoimmune Diseases: Increased 47,XXX in  
Systemic Lupus Erythematosus and Sjögren's Syndrome**

Ke Liu BS<sup>1,2</sup>, Bijl T. Kurien PhD<sup>3,4,5</sup>, Sarah L. Zimmerman PhD<sup>6</sup>, Kenneth M. Kaufman PhD<sup>1,7</sup>,  
Diana H. Taft PhD<sup>1</sup>, Leah C. Kottyan PhD<sup>1</sup>, Sara Lazaro MBA, MS<sup>1</sup>, Carrie A. Weaver MA<sup>1</sup>,  
John A. Ice PhD<sup>4</sup>, Adam J. Adler BS<sup>4,5</sup>, James Chodosh MD, MPH<sup>8</sup>, Lida Radfar DDS<sup>9</sup>, Astrid  
Rasmussen MD, PhD<sup>4</sup>, Donald U. Stone MD<sup>10</sup>, David M. Lewis DDS<sup>9</sup>, Shibo Li MD<sup>3</sup>, Kristi A.  
Koelsch PhD<sup>3,4</sup>, Ann Igoe MD<sup>3,4</sup>, Mitali Dave MD<sup>3</sup>, Jay Kumar BS<sup>3,4</sup>, Jacen S. Maier-Moore  
PhD<sup>3,4,5,11</sup>, Valerie M. Harris MS<sup>3,4</sup>, Rajaram Gopalakrishnan DDS<sup>12</sup>, Roland Jonsson DMD,  
PhD<sup>13,14</sup>, James A. Lessard MD<sup>15</sup>, Xianglan Lu PhD<sup>3</sup>, Jacques-Eric Gottenberg MD, PhD<sup>16</sup>,  
Juan-Manuel Anaya MD, PhD<sup>17</sup>, Deborah S. Cunninghame-Graham PhD<sup>18</sup>, Andrew J. W.  
Huang MD, MPH<sup>12</sup>, Michael T. Brennan DDS, MHS<sup>19</sup>, Pamela Hughes MD, PhD<sup>12</sup>, Gabor G.  
Illei MD<sup>20</sup>, Corinne Miceli-Richard MD, PhD<sup>21</sup>, Edward C. Keystone MD<sup>22</sup>, Vivian P. Bykerk  
MD<sup>23</sup>, Gideon Hirschfield PhD<sup>24</sup>, Gang Xie PhD<sup>25</sup>, Wan-Fai Ng PhD<sup>26</sup>, Gunnel Nordmark MD,  
PhD<sup>27</sup>, Per Eriksson PhD<sup>28</sup>, Roald Omdal MD, PhD<sup>29</sup>, Nelson L. Rhodus MPH, DDS<sup>30</sup>,  
Maureen Rischmueller MBBS<sup>31,32</sup>, Michael Rohrer MD<sup>12</sup>, Barbara M. Segal MD<sup>33</sup>, Timothy J.  
Vyse MD, PhD<sup>18</sup>, Marie Wahren-Herlenius MD, PhD<sup>34</sup>, Torsten Witte MD<sup>35</sup>, Xavier Mariette  
MD, PhD<sup>36</sup>, Bernardo Pons-Estel MD<sup>37</sup>, Marta E. Alarcon-Riquelme MD, PhD<sup>4,38</sup>, Joel M.  
Guthridge PhD<sup>3,4</sup>, Judith A. James MD, PhD<sup>3,4</sup>, Christopher J. Lessard PhD<sup>3,4</sup>, Jennifer A. Kelly  
MPH<sup>4</sup>, Susan D. Thompson PhD<sup>1</sup>, Patrick M. Gaffney MD<sup>4</sup>, Courtney G. Montgomery PhD<sup>4</sup>,  
Katherine A. Siminovitch MD<sup>25</sup>, Jeffrey C Edberg PhD<sup>39</sup>, Robert P Kimberly MD<sup>39</sup>, Graciela S.  
Alarcón MD, MPH<sup>39</sup>, Carl L. Langefeld PhD<sup>40</sup>, Gary S. Gilkeson MD<sup>41,42</sup>, Diane L. Kamen  
MD, MS<sup>42</sup>, Betty P. Tsao PhD<sup>43</sup>, W. Joseph McCune MD<sup>44</sup>, Jane E. Salmon MD<sup>45</sup>, Joan T.  
Merrill MD<sup>3</sup>, Michael H Weisman MD<sup>46</sup>, Daniel J Wallace MD<sup>46</sup>, Tammy O Utset MD<sup>47</sup>, Erwin

**X CHROMOSOME DOSE AND SEX BIAS IN AUTOIMMUNE DISEASES**  
**ar-15-0012 revised manuscript**

P. Bottinger MD<sup>48</sup>, Christopher I. Amos PhD<sup>49</sup>, Kathy L. Sivils PhD<sup>3,4</sup>, John B. Harley MD, PhD<sup>1,2,7,\*</sup>, R. Hal Scofield MD<sup>3,4,5,\*</sup>.

\*To whom co-correspondence should be addressed, these authors contributed equally to this study.

**Correspondence:**

R. Hal Scofield, MD

825 NE 13<sup>th</sup> Street, MS24

Oklahoma City, OK 73104 USA

Phone: +1 405 271 7061

Fax: +1 405 271 7063

E-mail: hal-scofield@omrf.ouhsc.edu

John B. Harley, MD, PhD

3333 Burnet Ave, ML 15012,

Cincinnati, Ohio 45229

Phone: +1 513 803 3665

Fax: +1 513 803 4652

Email: John.Harley@cchmc.org

**Affiliations:**

<sup>1</sup>Center for Autoimmune Genomics and Etiology (CAGE), Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, USA.

<sup>2</sup>College of Medicine, University of Cincinnati, Cincinnati, Ohio, USA.

<sup>3</sup>College of Medicine, University of Oklahoma Health Sciences Center, Oklahoma City, OK 73104, USA.

<sup>4</sup>Arthritis & Clinical Immunology Program, Oklahoma Medical Research Foundation, Oklahoma City, OK 73104, USA.

<sup>5</sup>U.S. Department of Veterans Affairs Medical Center, Oklahoma City, OK 73104, USA.

<sup>6</sup>Division of Human Genetics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA.

<sup>7</sup>U.S. Department of Veterans Affairs Medical Center, Cincinnati, Ohio, USA.

<sup>8</sup>Department of Ophthalmology, Howe Laboratory, Massachusetts Eye and Ear Infirmary, Harvard Medical School, Boston, Massachusetts, USA.

<sup>9</sup>Department of Oral Diagnosis and Radiology, College of Dentistry, University of Oklahoma Health Sciences Center, Oklahoma City, USA.

<sup>10</sup>Dean McGee Eye Institute and Department of Ophthalmology, University of Oklahoma College of Medicine, Oklahoma City, OK, USA.

<sup>11</sup>Department of Clinical Laboratory Sciences, University of Texas at El Paso, El Paso, TX 79968.

<sup>12</sup>Department of Developmental and Surgical Sciences, University of Minnesota, Minneapolis, MN, USA.

<sup>13</sup>Broegelmann Research Laboratory, Department of Clinical Science, University of Bergen, Bergen 5021, Norway.

<sup>14</sup>Department of Rheumatology, Haukeland University Hospital, Bergen 5021, Norway.

<sup>15</sup>Valley Bone & Joint Clinic, 3035 DeMers Avenue, Grand Forks, ND 58201, USA.

<sup>16</sup>Strasbourg University, Strasbourg, France.

<sup>17</sup>Center for Autoimmune Diseases Research (CREA), School of Medicine and Health Sciences, Universidad del Rosario, Bogota, Colombia.

<sup>18</sup>Division of Genetics and Molecular Medicine and Division of Immunology, Infection and Inflammatory Disease, King's College London, London.

<sup>19</sup>Department of Oral Medicine, Carolinas Medical Center, Charlotte, NC 28232, USA.

<sup>20</sup>Sjögren's Syndrome Clinic, National Institute of Dental and Craniofacial Research, Molecular Physiology and Therapeutics Branch, National Institutes of Health, Bethesda, MD 20892, USA.

<sup>21</sup>Department of Rheumatology, Université Paris-Sud, AP-HP, INSERM U1012, Le Kremlin-Bicêtre, France.

<sup>22</sup>Department of Medicine, Mount Sinai Hospital and University of Toronto, Toronto, Ontario.

<sup>23</sup>Hospital for Special Surgery, New York, USA.

<sup>24</sup>NIHR Biomedical Research Unit, University of Birmingham, Birmingham, UK.

<sup>25</sup>Samuel Lunenfeld and Toronto General Research Institutes, Departments of Medicine, Immunology and Molecular Genetics, University of Toronto, Toronto Ontario.

<sup>26</sup>Musculoskeletal Research Group, Institute of Cellular Medicine & NIHR Newcastle Biomedical Research Centre, Newcastle University, Newcastle upon Tyne, United Kingdom.

<sup>27</sup>Section of Rheumatology, Department of Medical Sciences and Science for Life Laboratory, Uppsala University, Uppsala, Sweden.

<sup>28</sup>Rheumatology, Department of Clinical and Experimental Medicine, Linköping University, Linköping, Sweden.

<sup>29</sup>Clinical Immunology Unit, Department of Internal Medicine, Stavanger University Hospital, Stavanger, Norway.

<sup>30</sup>Department of Diagnostic and Biological Sciences, School of Dentistry, University of Minnesota, Minneapolis, USA.

<sup>31</sup>Rheumatology Department, The Queen Elizabeth Hospital, Woodville South, SA 5011, Australia.

<sup>32</sup>Discipline of Medicine, University of Adelaide, Adelaide, SA 5000, Australia.

<sup>33</sup>Department of Medicine, University of Minnesota Medical School, Minneapolis, MN, USA.

<sup>34</sup>Rheumatology Unit, Department of Medicine, Karolinska Institutet, Stockholm, Sweden.

<sup>35</sup>Clinic for Immunology and Rheumatology, Hannover Medical School, 30625 Hannover, Germany.

<sup>36</sup>Department of Rheumatology, Université Paris-Sud, AP-HP, INSERM U1012, Le Kremlin-Bicêtre, France.

<sup>37</sup> Sanatorio Parque, Rosario, Argentina.

<sup>38</sup>Center Pfizer, University of Granada, Andalusian Government for Genomics and Oncological Research, PTS Granada, 18016, Spain

<sup>39</sup>Musculoskeletal and Autoimmunity Center, Division of Clinical Immunology and Rheumatology, The University of Alabama, Birmingham, AL.

<sup>40</sup>Department of Biostatistical Sciences, Division of Public Health Sciences, Wake Forest University, Winston-Salem, NC.

<sup>41</sup>Division of Rheumatology & Immunology, Medical University of South Carolina, Charleston, SC.

<sup>42</sup>Ralph H. Johnson VA Medical Center, Charleston, SC.

<sup>43</sup>Division of Rheumatology, Department of Medicine, UCLA School of Medicine.

<sup>44</sup>Division of Rheumatology, Department of Medicine, University of Michigan College of Medicine, Ann Arbor, Michigan.

<sup>45</sup>Division of Rheumatology, Hospital for Special Surgery and Weill Cornell Medical College, New York, NY.

<sup>46</sup>Division of Rheumatology, Cedars-Sinai Medical Center, Los Angeles, CA.

<sup>47</sup>University of Chicago Pritzker School of Medicine, Chicago, IL.

<sup>48</sup>Charles R. Bronfman Institute for personalized medicine, Mount Sinai Hospital, 1468 Madison Avenue, New York, NY 10029.

<sup>49</sup>Department of Community and Family Medicine, Geisel School of Medicine, Dartmouth College, Hanover, NH, USA.

**Competing interests:**

We declare that we have no conflict of interest.

**Funding:**

This work was funded by the USA National Institutes of Health (AR062755, RR029882, AI024717, AI031584 , AI062629 , AI063274, AI082714, AI083194, AR042460, AR48204, AR048940, AR049084, AR052125, AR053483, AR053734, AR056360, AR058959, AR62277, AR043814, AR065626, DE015223, DE018209, RR015577, RR020143, GM103510, GM104938 and HG006828), the Intramural Research Program of the National Institute of Dental and Craniofacial Research, the University of Oklahoma Health Sciences Center and its Clinical and Translational Science OCTSI Summer Scholar Program, the U.S. Department of Veterans Affairs (IMMA9 to RHS and JBH), the U.S. Department of Defense (PR094002), Alliance for

Lupus Research (JBH), Mary Kirkland Scholar (JAJ and JBH), the Strategic Research Program at Helse Bergen, The Western Norway Regional Health Authority, The Broegelmann Foundation, EvASSESS PHRC (Programme Hospitalier de Recherche Clinique) 2006 from the French Ministry of Health, the Swedish Rheumatism Foundation, Arthritis Australia, an unrestricted grant from Research to Prevent Blindness to the Dean McGee Eye Institute and the Department of Ophthalmology, University of Oklahoma College of Medicine, the Medical Research Council, UK (G0800629), DFG KFO 250 WI 1031/6-1 (TW), a Senior Scientific Investigator Award to James Chodosh from Research to Prevent Blindness, and the ISCIII (02558) through FEDER funds (MEAR).



**Abstract:**

***Objective:***

More than 80% of autoimmune disease is female dominant, but the mechanism for this female bias is poorly understood. We suspected an X chromosome dose effect and hypothesized that trisomy X (47,XXX , 1 in ~1,000 live female births) would be increased in female predominant diseases (e.g. systemic lupus erythematosus [SLE], primary Sjögren's syndrome [SS], primary biliary cirrhosis [PBC] and rheumatoid arthritis [RA]) compared to diseases without female predominance (sarcoidosis) and controls.

***Methods:***

We identified 47,XXX subjects using aggregate data from single nucleotide polymorphism (SNP) arrays and confirmed, when possible, by fluorescent in situ hybridization (FISH) or quantitative polymerase chain reaction (q-PCR).

***Results:***

We found 47,XXX in seven of 2,826 SLE and three of 1,033 SS female patients, but only in two of the 7,074 female controls ( $p=0.003$ , OR=8.78, 95% CI: 1.67-86.79 and  $p=0.02$ , OR=10.29, 95% CI: 1.18-123.47; respectively). One 47,XXX subject was present for ~404 SLE women and ~344 SS women. 47,XXX was present in excess among SLE and SS subjects.

***Conclusion:***

The estimated prevalence of SLE and SS in women with 47,XXX was respectively ~2.5 and ~2.9 times higher than in 46,XX women and ~25 and ~41 times higher than in 46,XY men. No statistically significant increase of 47,XXX was observed in other female-biased diseases (PBC or RA), supporting the idea of multiple pathways to sex bias in autoimmunity.

## **Introduction**

Autoimmune disease affects 5-10% of the global population. Female preponderance is highlighted by the fact that almost 80% of autoimmune patients are female(1). Systemic lupus erythematosus (SLE), primary Sjögren's syndrome (SS), primary biliary cirrhosis (PBC), rheumatoid arthritis (RA) all have a female sex bias, though the ratios vary from fourteen women to one man in SS to only two to six women to one man in RA(2-5).

Several hypotheses have been proposed to explain the female preponderance in autoimmunity, such as sex hormones(6) and X monosomy(7, 8), but origins remain unknown. We have previously shown that Klinefelter's syndrome (47,XXY) is found in excess among men with SLE, and the calculated prevalence of SLE among Klinefelter's syndrome is similar to the prevalence in women (46,XX)(9). Data for SLE in four core genotype mice show that the disease is associated with the number of X chromosomes, and not phenotypic sex (10, 11). In aggregate these data are consistent with the gene dose of chromosome X being a major determining factor for the female predominance of SLE in man and mouse.

In order to further test this X chromosome dose hypothesis, we asked whether having a third X chromosome increases the risk of developing SLE or other autoimmune diseases. Trisomy X (47,XXX) is present in approximately 1 out of every 1000 live-born girls(12, 13). There are no consistently recognized abnormalities in sex hormone levels, sexual development, fertility, or pregnancy(14), though premature ovarian failure has been associated with carriage of 47,XXX(15-17). One suggested phenotype difference is that 47,XXX women may have more anxiety and shyness and lower self-esteem(18). In fact, 47,XXX is usually unrecognized, with

over 90% of the patients undiagnosed unless patients are screened for chromosomal abnormalities for other reasons(14). Herein, we show that 47,XXX is enriched among patients with SLE and SS, but not enriched in patients with PBC, RA or the non-autoimmune disease sarcoidosis.

## **Materials and Methods**

### ***Subjects and Genotyping Methods***

We used genotyping records of subjects enrolled in multiple different genetic studies (some published) for this work. We excluded all male subjects from our study. To be included as a case in this study, subjects need to have a confirmed diagnosis of a disease of interest: SLE, SS, PBC, RA or sarcoidosis. For controls, we used non-affected SLE family controls, non-auto-inflammatory subjects, population controls and non-sarcoidosis controls. A summary of recruitment and inclusion information for each cohort used in this study is available in the supplementary information (**Supplementary table 1a, 1b**). Institutional review boards (IRB) at each site provided approval for this study, and individual informed consent was obtained from all participants.

All patients with SLE met the American College of Rheumatology (ACR) classification criteria for systemic lupus erythematosus(19). Subjects with SS met the 2002 American-European Consensus Group (AECG) classification criteria for primary Sjögren's syndrome (20). PBC patients met the American Association for the Study of Liver Diseases Diagnostic Criteria for PBC(21). Participants with RA met 1987 American College of Rheumatology criteria for the classification of RA (22). Sarcoidosis patients had a) tissue confirmation of granuloma, or b)

chest radiographic evidence of bilateral symmetrical hilar adenopathy with either a history of erythema nodosum or at least two years observation during which time no other disease was found to explain radiographic abnormalities(23-26) (**Supplementary table 1a**).

Controls include first degree family members (sisters and mothers) of SLE patients, all of whom were shown not to have SLE (27). Other controls are healthy individual with no auto-inflammatory diseases from three cohorts (28-30). Population controls of European decedents recruited globally were also added (31). Another population control group consisted of self-identified African Americans receiving care at Mt. Sinai Hospital in New York. The last control cohorts were recruited as part of the sarcoidosis studies, and were without sarcoidosis (23-26) (**Supplementary table 1b**).

Genotyping records are from various platforms: ImmunoChip [196,524 single nucleotide polymorphisms (SNPs), Illumina], HumanHap370 Beadchip (300,000 SNPs, Illumina). Omni1 chip (1,134,514 SNPs, Illumina), OmniExpress chip (741,000 SNPs, Illumina) or Omni5-Quad chip (4,301,331 SNPs, Illumina). A summary of genotyping platforms for each cohort used in this study is available in the supplementary information (**Supplementary table 1a, 1b**).

#### ***47,XXX identification***

X chromosome copy number variations were found by visually inspecting B allele frequency and LogR ratio (LRR) plot of each subject's X chromosome for abnormalities using Genome Studio (Illumina), then mean logR ratio was calculated to confirm gain of copy number. In the B allele frequency plot, fluorescence intensity of the B allele of any SNP is plotted over the total

fluorescence for that SNP in a given individual. Therefore, at a SNP with BB homozygosity the result is 100%, with AA homozygosity the result is 0%, and with AB heterozygosity the result is 50%. In the LRR plot, the fluorescence intensity of total allele A and B of the given sample is normalized by the reference genome to differentiate between copy number gains and losses.

Normal females (46,XX) were identified by the heterozygous "three band" pattern in the B allele frequency plot, which corresponds to the 0%, 50%, and 100% SNP frequencies (i.e. AA, AB, BB). The mean logR ratio, shown by a red line at 0.0 in the LRR plot, indicates there are two copies of the X chromosome (**Figure 1**).

Trisomy X (47,XXX) subjects were identified by the "four band" pattern in the B allele frequency plot, which corresponds to 0%, 33%, 66%, and 100% SNP frequencies (i.e. AAA, AAB, ABB, BBB) with the accompanying mean LRR near 0.2 (**Figure 1**).

### ***Fluorescence in situ hybridization (FISH)***

When cells were available, samples with chromosome abnormalities identified by B allele frequency and LRR plots were validated by FISH, as previously reported(32). For subjects that had frozen peripheral blood mononuclear cells (PBMCs), we used commercial FISH probes that recognize the centromeres of chromosomes X and Y to confirm aneuploidies of chromosome X (alpha-satellite repeats DXZ1 & DYZ3, PID# KI-20030, Veridex). Fifty nuclei were scored for each subject. For subjects that only had Epstein-Barr virus transformed B lymphocyte cell lines (LCLs or lymphoblastoid cell lines) available, we used commercial FISH probes that recognize the X centromeres, Yp and chromosome 18 (DXZ1, DYZ3 and D18Z1, Vysis) to confirm

aneuploidies of chromosome X. Chromosome 18 probe signals served as controls to establish that the cell lines tested had a normal diploid genome. Two hundred nuclei were scored for each subject. Possible mosaicism was assessed according to the percentage of 45,X, 46,XX, or 47,XXX cells enumerated (**Figure 2**).

### ***Polymerase Chain Reaction (PCR)***

47,XXX samples having only DNA available were confirmed by PCR using TaqMan® copy variation assays (Life technology). Copy number of a gene on the X chromosome (YIPF6) was normalized to TaqMan® copy number reference assay, RNase P, on chromosome 14. Real time PCR was performed on an Applied Biosystems 7500 thermocycler. Data were analyzed using CopyCaller v2.0 software (**Figure 3**).

### ***Quality control and statistical analysis***

Subjects that had a call frequency lower than 0.90, genetically XY individuals self-reported as females, subjects whose samples indicated mixing of samples from multiple individuals or duplicated samples were removed.

The enrichment of 47,XXX in SLE, SS, PBC, RA subjects versus control subjects were tested by two-tailed Fisher's exact test using R (Version 3.2.2) (**Table 1**).

The prevalence of SLE and SS among women with 47,XXX was estimated using Bayes' theorem. Below are the Bayes' theorem ( $P [B/A] = (P [A/B]*P [B])/P [A]$ ) calculations to estimate the prevalence of SLE and Sjögren's syndrome in females with 47,XXX.

A = the frequency of trisomy X, B = the frequency of SLE or SS in females, P [A/B] represents frequency of individuals with 47,XXX in SLE or SS, and P [B/A] would indicate the probability of SLE or SS within the group with trisomy X.

<u>Condition</u>	<u>Prevalences used in the Bayes' theorem calculations</u>
1) SLE	= 1 in 1,000 women in the USA(2)
2) SS	= 1 in 200 women in the USA(33)
3) 47,XXX	= 1 in 1,000 women(34)
4) 47,XXX with SLE	= 1 in 404 SLE women (data from this study)
5) 47,XXX with SS	= 1 in 344 SS women (data from this study)

The Bayes' theorem calculations:

Prevalence of *SLE in females with 47, XXX*

$$= [Freq (47, XXX in SLE) * Freq (SLE in females)] / Freq (47, XXX in females)$$

Prevalence of *SS in females with 47, XXX*

$$= [Freq (47, XXX in SS) * Freq (SS in female)] / Freq (47, XXX in females)$$

Population stratification was performed using SNP & Variation Suite software (Golden Helix) for SLE cases and controls. European ancestry and African ancestry were segregated within 3 s.d. of the mean of the first 2 principal components. Non-European or African ancestry subjects were dropped as well as the ones that their self-identified races did not match the principal component analysis results. Logistic regression modeling in R (Version 3.0.0) using glm function was performed to adjust odds ratio values for ancestry.

To confirm our results were not due to biases in the control groups, we also performed tests comparing 47,XXX enrichment in our case subjects to different control groups. The enrichment of 47,XXX in SLE versus non-affected SLE family controls was tested by two-tailed Fisher's exact test using R (Version 3.2.2). The enrichment of 47,XXX in SS versus healthy non-auto-inflammatory controls was also tested by two-tailed Fisher's exact test using R (Version 3.2.2). The enrichment of SLE and SS subjects versus a) controls without non-sarcoidosis subjects, b) controls without population controls and c) controls without non-sarcoidosis and population controls were also tested by two-tailed Fisher's exact test using R (Version 3.2.2). Finally, the same analysis was used to test the enrichment of 47,XXX in SLE, SS, PBC, RA subjects versus unselected newborn infants controls as reported by Jacob et.al(34).

## **Results**

### ***Subjects***

After quality control, we had 2826 SLE, 1033 SS, 1118 PBC, 1710 RA, 939 sarcoidosis and 7074 control subjects for our study. For controls, we had 2090 non-affected first degree family member of SLE patients, 1684 non-auto-inflammatory healthy control subjects, 2680 population control subjects and 620 non-sarcoidosis control subjects (**Supplementary table 1a, 1b**). There is no difference in prevalence of 47,XXX between SLE family control versus the rest controls by Fisher exact test ( $p=1$ , data not shown). A risk with the population controls is that some subjects might have the diseases of interest because this group represents a general population and was not specifically screened for the diseases of interest to this study. The non-sarcoidosis controls (n=620) may also contain subjects with diseases of interest (other than sarcoidosis) but the



smaller number of subjects makes it less likely to contain SLE/SS/PBC/RA patients. Since the non-sarcoidosis control and population control each contains one of the two 47,XXX in the controls, we decided to retain them for modeling purpose. This decision potentially biases our results towards the null hypothesis. We therefore considered all three control cohorts in aggregate for our primary analysis.

***47,XXX enrichment in systemic lupus erythematosus and primary Sjögren's syndrome***

Using single nucleotide polymorphism typing of the X chromosome, we found 47,XXX in seven of 2,826 women with SLE and in three of 1,033 women with SS, but only two in the 7,074 controls ( $p=0.003$ ,  $OR=8.78$ , 95% CI: 1.67-86.79 and  $p=0.02$ ,  $OR=10.29$ , 95% CI: 1.18-123.47; respectively; **Table 1, Fig. 1**). Five of the seven 47,XXX subjects with SLE were validated by FISH (**Fig. 2**). We validated one using frozen PBMCs and another four using lymphoblastoid cell lines (LCLs). Four subjects had 100% of 47,XXX cells, one had 98.5% of 47,XXX cells and 1.5% of 48,XXXX cells, which led us to consider this subject with the other 47,XXX subjects. We validated the two other 47,XXX subjects with SLE and the three with primary SS using q-PCR, which demonstrated a three X chromosome complement in all five DNA samples (**Fig. 3**). No DNA or cells were available from the two 47,XXX controls, but the complete fidelity between the genotyping screening and more traditional method supports the likelihood of accuracy in these samples.

Thus, we find that 1 in 404 (95% CI: 196-1004) women with SLE and 1 in 344 (95% CI: 115-1620) with SS have 47,XXX, ~2.5 and ~2.9 times higher, respectively, than the published population prevalence of 47,XXX which is ~1/1000(12, 13).

Controls in this study consisted of non-affected SLE family members, non-auto-inflammatory subjects, population controls, and non-sarcoidosis controls. The best controls for the SLE patients are the 2090 non-affected SLE family members. However, there are no 47,XXX participants found in this control cohort. There is a significant difference in 47,XXX enrichment when comparing SLE patients only to the family control cohort(**Supplementary table 3**). The non-affected SLE family members could possibly contain subjects with SS, so we tested 47,XXX enrichment in SS with only the 1684 healthy non-auto-inflammatory controls and the results remained significant (**Supplementary table 4**). The 2680 population controls and 620 non-sarcoidosis controls could potentially contain subjects with SLE or SS. We tested 47,XXX enrichment in SLE and SS with the removal of either or both of these control datasets and the results remained significant(**Supplementary table 5a, 5b, 5c**). We also compared 47,XXX in SLE and SS with a published unselected new born infants 47,XXX screening datasets (n=20,790) (34). Since a reported 47,XXX phenotype is increased anxiety and shyness combined with lower self-esteem(18), we feared 47,XXX participants potential may be less willing to participate in research studies, although this should equally affect case and control recruitment. Using the unselected new born infants' dataset should have no recruitment bias but a potential bias in cases remains. We still have a significant enrichment of 47,XXX in SLE ( $p=0.036$ , OR=2.58, 95% CI: 0.92-6.36) and the same trend for 47,XXX enrichment in SS ( $p=0.093$ , OR=3.02, 95% CI: 0.57-10.22) (**Supplementary table 6**). We are therefore inclined to conclude that the comparison with the combination of all control cohorts is a fair representation of the enrichment in SLE and SS. The loss of significance in SS maybe due to small sample size.

Using the published population prevalence of 47,XXX in the population (~1/1000) and Bayes' theorem(9, 12, 13), we predict that SLE exists in 1 of 404 women with 47,XXX and SS in 1 of 69 women with 47,XXX. This calculation assumes that SLE is present in 1/1,000, and SS in 1/200 women in the USA. We estimate the prevalence of SLE and SS with 47,XXX is ~2.5 times higher than that with 46,XX and ~25 and ~41 times higher than 46,XY, respectively (**Table 3**).

***47,XXX risk is independent of ancestry***

Individuals of African ancestry have a higher risk for SLE compared to those of European ancestry(35), and we observed a majority of the 47,XXX subjects self-identified as of African ancestry. While there are no studies on 47,XXX prevalence by ancestry, there is a study demonstrating 47,XXY differences by ancestry(34). Therefore, we asked whether the SLE risk associated with 47,XXX is dependent on ancestry. After population stratification, 375 samples were categorized as non-European/African ancestries and therefore excluded. An additional 144 samples were dropped because of a self-identified race mismatch with the principal component analysis, giving us a final of 1592 subjects of African ancestry (AA) and 1083 subjects of European ancestry (EA) with SLE, along with 3755 AA and 2951 EA control subjects. Among the seven 47,XXX women with SLE, five self-identified as being of African ancestry, and two of European ancestry. The two control 47,XXX women self-identified as of African ancestry. After principal component analysis for population stratification, one African ancestry SLE 47,XXX subject was removed as an admixed outlier (**Table 2**). We performed logistic regression analysis of SLE status with 47,XXX and ancestry as independent variables. The ancestry adjusted OR was 7.36 with p=0.01, similar to the unadjusted model (**Supplementary Table 2**). We also modeled using the interaction term between ancestry and 47,XXX to test for effect modification,

but the interaction term was not significant. So we conclude that the odds of 47,XXX in SLE is independent of African ancestry in our data.

In addition, there was an increased incidence of 47,XXX among the SS patients. The SS patient cohort was almost entirely comprised of individuals of European ancestry (>95%); all three 47,XXX SS cases self-identified as being of European ancestry. This meant we were unable to adjust for ancestry effects in SS.

#### ***47, XXX in other autoimmune diseases***

We found 47,XXX in 1 of 1,118 women with PBC, 1 of 1,710 with RA, and none of 939 with sarcoidosis ( $p>0.05$ , **Table 1**). The 47,XXX subject with PBC was validated to have three X chromosomes with q-PCR (**Figure 3**). No DNA or cells were available for the 47,XXX subject with RA. There was no statistically significant increase in prevalence of 47,XXX seen in PBC, RA or sarcoidosis compared to our controls. The comparison of 47,XXX in PBC, RA and sarcoidosis to the unselected newborn infants dataset also found no significant difference (**Supplementary table 6**).

#### **Discussion**

We showed that 47,XXX is more frequent in patients with SLE and SS. Indeed, this and previous work show that the risk of SLE and SS increases with each additional X chromosome (**Table 3**). The ~25-fold increase in predicted prevalence for SLE and ~41-fold increase for SS observed in 47,XXX women compared to 46,XY men reveals an effect on risk exceeding by many fold that of all other known genetic risk factors for SLE or SS. The importance of these findings is not for the few individuals with 47,XXX, but rather lies in the fact that rare events and phenotypes

reveal insights into the mechanism for the general disease circumstances. Everyone has an X chromosome, and increased 47,XXX among women with either SLE or SS informs the potential mechanism underpinning the disparate risk of these diseases found for men and women with a normal sex chromosome complement. Because sexual development and sex hormones are normal in 47,XXX women, these data suggest that the number of X chromosomes is a key factor imparting the 10-fold risk difference between men and women.

Because 47,XXX is a rare event, in order to get a sufficiently large sample size, we included non-affected SLE family controls, non-auto-inflammatory subjects, population controls and non-sarcoidosis controls in our study. The 2090 non-affected SLE family controls may contain subjects with SS, we showed that when comparing 47,XXX enrichment in SS only with the healthy non-auto-inflammatory controls, the result is still significant (**Supplementary table 4**). The 2680 population control and 620 non-sarcoidosis controls may contain subjects with SLE or SS. We showed that even when removing these control groups, our 47,XXX enrichment in SLE and will still be significant if not more so (**Supplementary table 5a,5b,5c**). One possible phenotype associated with 47,XXX is anxiety, shyness and low self-esteem(18), a potential source of bias as 47,XXX subjects may be less willing to participate in research studies. This raises the question of whether 47,XXX would be less willing to participate in our study. We think this personality trait of 47,XXX women would lead to a non-differential error that would equally affect both case and control group recruitment. The majority (90%) of the 47,XXX are unrecognized(14), which means they only rarely exhibit a clear phenotype or that any phenotype is typically mild. We also compared our rate of 47,XXX in SLE, SS, RA, PBC, and sarcoidosis with a published unselected new born infants 47,XXX screening dataset(34). The newborn infant

dataset has no bias in recruitment since consecutive births were studied (**Supplementary table 6**). Even in this case, we still have a significant enrichment of 47,XXX in SLE and a trend for 47,XXX enrichment in SS, which supports our hypothesis of X chromosome does effect in SLE and SS.

We did not find excess 47,XXX among women with the other autoimmune diseases studied here (PBC and RA) compared to our controls or the unselected newborn infant dataset (**Table 1, Supplementary table 6**). While it is possible that we had inadequate sample size for PBC and RA, the relationship of Turner's syndrome (45,X) to autoimmunity also suggests heterogeneity in the mechanisms of the sex bias found in autoimmune diseases. Autoimmune thyroiditis, type 1 diabetes mellitus, and celiac disease are found in excess among Turner's syndrome (45,X) patients(36). On the other hand, there is no evidence that prevalence of SLE or SS are increased in Turner's syndrome. While rarer than the Turner's-associated diseases, the paucity of Turner's syndrome among SLE patients (37, 38) suggests that SLE risk in these patients is more similar to 46,XY males rather than to 46,XX females. Moreover, PBC, autoimmune thyroid disease (AITD) and systemic sclerosis have increased acquired X-monosomy but SLE does not(7, 8, 39), also suggesting existence of multiple mechanisms for sex bias in autoimmunity.

If SLE and 47,XXX both occur at 1 in ~1000 and were independent, then only 1 in a million women would have both SLE and 47,XXX. Our data suggest that both are coincident at 1 in ~40,000 women. The SLE families provide an interesting insight into how these risks may be modified. There are 7 women with SLE and 47,XXX in our sample of 2826 SLE cases, but no primary female relatives of SLE patients with 47,XXX among the 2090 subjects we examined.

That this difference is significant ( $p=0.02$ ) suggests that 47,XXX is a strong contributor to SLE when present within the family milieu of other genetics and environmental factors that establish risk for SLE (**Supplementary table 3**).

Sex hormones have long been considered a candidate mechanism of gender bias. We previously reported Klinefelter's men (47,XXY) have a similar risk of SLE as women (46,XX) (9), while in the present study, 47,XXX women, who do not produce abnormal sex hormones(14, 40), have excess risk for SLE and SS compared to 46,XX women. Furthermore, there is no difference in the levels of sex hormones between SLE patients at disease onset and healthy controls(41). Other data suggest that sex hormone levels in men with SLE are abnormal, but similar to those found in men with other non-female-biased chronic illnesses(42). The female bias in SLE also exists before puberty with girls being five times more likely to develop SLE than boys(43). Similarly, the sex bias in SLE is unaffected by the lower estrogen levels after menopause, and is not statistically different from those with disease onset in the reproductive years(44). These observations suggest that X chromosome dose contributing to sex bias in SLE is independent of hormonal effects. X chromosome dose effect predisposes women to greater susceptibility to SLE and SS.

Perhaps the observation of a global reduction in methylation in SLE by Richardson and colleagues in activated T cells of SLE patients (45) offers possible explanations for a gene dose hypothesis. Each additional X chromosome, which is normally inactivated by methylation, could exert an effect if methylation is reversed or if additional X chromosomes affected methylation-related regulation of autosomal genes. Alternatively, a gene or gene(s) that escapes X

inactivation (46), producing a gene-dose effect could be an explanation. Support for above mechanisms has not yet been developed.

To conclude, the available data strongly support a genetic cause of the powerful female bias in SLE and SS being located on the X chromosome with a mechanism that appears to be at least partially independent of circulating sex hormones. Our new data demonstrating increased risk of SLE and SS among 47,XXX women indicate that the female bias may be explained, at least in part, by an X chromosome gene dose effect. Identifying the specific properties of the X chromosome responsible for the increase in SLE and SS risk will be critical for advancing the understanding of these disorders.



## **References**

1. Eaton WW, Rose NR, Kalaydjian A, Pedersen MG, Mortensen PB. Epidemiology of autoimmune diseases in Denmark. *Journal of autoimmunity*. 2007;29(1):1-9.
2. Helmick CG, Felson DT, Lawrence RC, Gabriel S, Hirsch R, Kwoh CK, et al. Estimates of the prevalence of arthritis and other rheumatic conditions in the United States. Part I. Arthritis and rheumatism. 2008;58(1):15-25.
3. Selmi C, Bowlus CL, Gershwin ME, Coppel RL. Primary biliary cirrhosis. *Lancet*. 2011;377(9777):1600-9.
4. Gabriel SE. The epidemiology of rheumatoid arthritis. *Rheum Dis Clin North Am*. 2001;27(2):269-81.
5. Massardo L, Pons-Estel BA, Wojdyla D, Cardiel MH, Galarza-Maldonado CM, Sacnun MP, et al. Early rheumatoid arthritis in Latin America: low socioeconomic status related to high disease activity at baseline. *Arthritis care & research*. 2012;64(8):1135-43.
6. Quintero OL, Amador-Patarroyo MJ, Montoya-Ortiz G, Rojas-Villarraga A, Anaya JM. Autoimmune disease and gender: plausible mechanisms for the female predominance of autoimmunity. *Journal of autoimmunity*. 2012;38(2-3):J109-19.
7. Invernizzi P, Miozzo M, Battezzati PM, Bianchi I, Grati FR, Simoni G, et al. Frequency of monosomy X in women with primary biliary cirrhosis. *Lancet*. 2004;363(9408):533-5.
8. Invernizzi P, Miozzo M, Selmi C, Persani L, Battezzati PM, Zuin M, et al. X chromosome monosomy: a common mechanism for autoimmune diseases. *Journal of immunology (Baltimore, Md : 1950)*. 2005;175(1):575-8.
9. Scofield RH, Bruner GR, Namjou B, Kimberly RP, Ramsey-Goldman R, Petri M, et al. Klinefelter's syndrome (47,XXY) in male systemic lupus erythematosus patients: support for the

notion of a gene-dose effect from the X chromosome. Arthritis and rheumatism.

2008;58(8):2511-7.

10. Smith-Bouvier DL, Divekar AA, Sasidhar M, Du S, Tiwari-Woodruff SK, King JK, et al.

A role for sex chromosome complement in the female bias in autoimmune disease. *The Journal of experimental medicine*. 2008;205(5):1099-108.

11. Sasidhar MV, Itoh N, Gold SM, Lawson GW, Voskuhl RR. The XX sex chromosome complement in mice is associated with increased spontaneous lupus compared with XY. *Ann Rheum Dis*. 2012;71(8):1418-22.

12. Nielsen J, Wohler M. Chromosome abnormalities found among 34,910 newborn children: results from a 13-year incidence study in Arhus, Denmark. *Hum Genet*. 1991;87(1):81-3.

13. Jacobs PA, Browne C, Gregson N, Joyce C, White H. Estimates of the frequency of chromosome abnormalities detectable in unselected newborns using moderate levels of banding. *J Med Genet*. 1992;29(2):103-8.

14. Tartaglia NR, Howell S, Sutherland A, Wilson R, Wilson L. A review of trisomy X (47,XXX). *Orphanet J Rare Dis*. 2010;5:8.

15. Devi A, Benn PA. X-chromosome abnormalities in women with premature ovarian failure. *J Reprod Med*. 1999;44(4):321-4.

16. Baronchelli S, Conconi D, Panzeri E, Bentivegna A, Redaelli S, Lissoni S, et al. Cytogenetics of premature ovarian failure: an investigation on 269 affected women. *J Biomed Biotechnol*. 2011;2011:370195.

17. Jiao X, Qin C, Li J, Qin Y, Gao X, Zhang B, et al. Cytogenetic analysis of 531 Chinese women with premature ovarian failure. *Hum Reprod*. 2012;27(7):2201-7.

18. Otter M, Schrander-Stumpel CT, Curfs LM. Triple X syndrome: a review of the literature. *European journal of human genetics : EJHG*. 2010;18(3):265-71.
19. Tan EM, Cohen AS, Fries JF, Masi AT, McShane DJ, Rothfield NF, et al. The 1982 revised criteria for the classification of systemic lupus erythematosus. *Arthritis and rheumatism*. 1982;25(11):1271-7.
20. Shiboski SC, Shiboski CH, Criswell L, Baer A, Challacombe S, Lanfranchi H, et al. American College of Rheumatology classification criteria for Sjogren's syndrome: a data-driven, expert consensus approach in the Sjogren's International Collaborative Clinical Alliance cohort. *Arthritis care & research*. 2012;64(4):475-87.
21. Lindor KD, Gershwin ME, Poupon R, Kaplan M, Bergasa NV, Heathcote EJ. Primary biliary cirrhosis. *Hepatology*. 2009;50(1):291-308.
22. Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis and rheumatism*. 1988;31(3):315-24.
23. Design of a case control etiologic study of sarcoidosis (ACCESS). ACCESS Research Group. *Journal of clinical epidemiology*. 1999;52(12):1173-86.
24. Rybicki BA, Hirst K, Iyengar SK, Barnard JG, Judson MA, Rose CS, et al. A sarcoidosis genetic linkage consortium: the sarcoidosis genetic analysis (SAGA) study. *Sarcoidosis, vasculitis, and diffuse lung diseases : official journal of WASOG / World Association of Sarcoidosis and Other Granulomatous Disorders*. 2005;22(2):115-22.
25. Iannuzzi MC, Maliarik MJ, Poisson LM, Rybicki BA. Sarcoidosis susceptibility and resistance HLA-DQB1 alleles in African Americans. *American journal of respiratory and critical care medicine*. 2003;167(9):1225-31.

26. Adrianto I, Lin CP, Hale JJ, Levin AM, Datta I, Parker R, et al. Genome-wide association study of African and European Americans implicates multiple shared and ethnic specific loci in sarcoidosis susceptibility. *PloS one*. 2012;7(8):e43907.
27. Rasmussen A, Sevier S, Kelly JA, Glenn SB, Aberle T, Cooney CM, et al. The lupus family registry and repository. *Rheumatology (Oxford, England)*. 2011;50(1):47-59.
28. Hollenbach JA, Thompson SD, Bugawan TL, Ryan M, Sudman M, Marion M, et al. Juvenile idiopathic arthritis and HLA class I and class II interactions and age-at-onset effects. *Arthritis and rheumatism*. 2010;62(6):1781-91.
29. Hinks A, Cobb J, Marion MC, Prahalad S, Sudman M, Bowes J, et al. Dense genotyping of immune-related disease regions identifies 14 new susceptibility loci for juvenile idiopathic arthritis. *Nature genetics*. 2013;45(6):664-9.
30. Gregersen PK, Amos CI, Lee AT, Lu Y, Remmers EF, Kastner DL, et al. REL, encoding a member of the NF-kappaB family of transcription factors, is a newly defined risk locus for rheumatoid arthritis. *Nature genetics*. 2009;41(7):820-3.
31. Lessard CJ, Li H, Adrianto I, Ice JA, Rasmussen A, Grundahl KM, et al. Variants at multiple loci implicated in both innate and adaptive immune responses are associated with Sjogren's syndrome. *Nature genetics*. 2013;45(11):1284-92.
32. Dillon SP, Kurien BT, Li S, Bruner GR, Kaufman KM, Harley JB, et al. Sex chromosome aneuploidies among men with systemic lupus erythematosus. *Journal of autoimmunity*. 2012;38(2-3):J129-34.
33. Vitali C, Bombardieri S, Jonsson R, Moutsopoulos HM, Alexander EL, Carsons SE, et al. Classification criteria for Sjogren's syndrome: a revised version of the European criteria proposed by the American-European Consensus Group. *Ann Rheum Dis*. 2002;61(6):554-8.

34. Jacobs PA. The incidence and etiology of sex chromosome abnormalities in man. Birth defects original article series. 1979;15(1):3-14.
35. Danchenko N, Satia JA, Anthony MS. Epidemiology of systemic lupus erythematosus: a comparison of worldwide disease burden. *Lupus*. 2006;15(5):308-18.
36. Bonamico M, Pasquino AM, Mariani P, Danesi HM, Culasso F, Mazzanti L, et al. Prevalence and clinical picture of celiac disease in Turner syndrome. *J Clin Endocrinol Metab*. 2002;87(12):5495-8.
37. Cooney CM, Bruner GR, Aberle T, Namjou-Khales B, Myers LK, Feo L, et al. 46,X,del(X)(q13) Turner's syndrome women with systemic lupus erythematosus in a pedigree multiplex for SLE. *Genes and immunity*. 2009;10(5):478-81.
38. Takegami T, Nakao K, Nagayama Y, Fujita T, Hoshino T, Tsunematsu T, et al. [A case of SLE associated with Turner's syndrome of 45, XO/46, XXq+ mosaicism (author's transl)]. *Nihon Naika Gakkai Zasshi*. 1980;69(7):861-6.
39. Invernizzi P, Miozzo M, Oertelt-Prigione S, Meroni PL, Persani L, Selmi C, et al. X monosomy in female systemic lupus erythematosus. *Ann N Y Acad Sci*. 2007;1110:84-91.
40. Barr ML, Sergovich FR, Carr DH, Saver EL. The triplo-X female: an appraisal based on a study of 12 cases and a review of the literature. *Can Med Assoc J*. 1969;101(5):247-58.
41. Chang DM, Chang CC, Kuo SY, Chu SJ, Chang ML. Hormonal profiles and immunological studies of male lupus in Taiwan. *Clin Rheumatol*. 1999;18(2):158-62.
42. Mackworth-Young CG, Parke AL, Morley KD, Fotherby K, Hughes GR. Sex hormones in male patients with systemic lupus erythematosus: a comparison with other disease groups. *Eur J Rheumatol Inflamm*. 1983;6(3):228-32.

43. Kamphuis S, Silverman ED. Prevalence and burden of pediatric-onset systemic lupus erythematosus. *Nat Rev Rheumatol*. 2010;6(9):538-46.
44. Lalani S, Pope J, de Leon F, Peschken C. Clinical features and prognosis of late-onset systemic lupus erythematosus: results from the 1000 faces of lupus study. *J Rheumatol*. 2010;37(1):38-44.
45. Lu Q, Wu A, Tesmer L, Ray D, Yousif N, Richardson B. Demethylation of CD40LG on the inactive X in T cells from women with lupus. *Journal of immunology (Baltimore, Md : 1950)*. 2007;179(9):6352-8.
46. Carrel L, Willard HF. X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nature*. 2005;434(7031):400-4.
47. Hirschfield GM, Liu X, Xu C, Lu Y, Xie G, Gu X, et al. Primary biliary cirrhosis associated with HLA, IL12A, and IL12RB2 variants. *The New England journal of medicine*. 2009;360(24):2544-55.
48. Juran BD, Hirschfield GM, Invernizzi P, Atkinson EJ, Li Y, Xie G, et al. Immunochip analyses identify a novel risk locus for primary biliary cirrhosis at 13q14, multiple independent associations at four established risk loci and epistasis between 1p31 and 7q32 risk variants. *Human molecular genetics*. 2012;21(23):5209-21.
49. Lopez Herraiz D, Martinez-Bueno M, Riba L, Garcia de la Torre I, Sacnun M, Goni M, et al. Rheumatoid arthritis in Latin Americans enriched for Amerindian ancestry is associated with loci in chromosomes 1, 12, and 13, and the HLA class II region. *Arthritis and rheumatism*. 2013;65(6):1457-67.

50. Karlson EW, Sanchez-Guerrero J, Wright EA, Lew RA, Daltroy LH, Katz JN, et al. A connective tissue disease screening questionnaire for population studies. *Annals of epidemiology*. 1995;5(4):297-302.

**Tables**

**Table 1. 47,XXX incidence.** 47,XXX in systemic lupus erythematosus (SLE), primary Sjögren's syndrome (SS), primary biliary cirrhosis (PBC), rheumatoid arthritis (RA), sarcoidosis and controls. Each disease state is compared to the control group. P-value, OR with 95% CI were calculated by Fisher exact test.

<b>Disease</b>	<b>F/M ratio</b>	<b>Sample Size</b>	<b>47,XXX</b>	<b>Odds Ratio (95%CI)</b>	<b>p value</b>	<b>Incidence</b>
Systemic lupus erythematosus	10:1	2826	7	8.78 (1.67-86.79)	0.003	1/404
Sjögren's syndrome	14:1	1033	3	10.29 (1.18-123.47)	0.02	1/344
Primary Biliary Cirrhosis	10:1	1118	1	3.16 (0.05-60.89)	NS	1/1118
Rheumatoid Arthritis	2~3:1	1710	1	2.07 (0.03-39.75)	NS	1/1710
Sarcoidosis	-	939	0	-	-	-
Controls	-	7074	2	-	-	1/3537



**Table 2. Numbers of 47,XXX by ancestry in SLE cases and controls.**

<b>Disease</b>	<b>Race</b>	<b>47,XXX</b>	<b>Sample Size</b>	<b>Total</b>
Systemic lupus erythematosus	AA	4	1592	2675
	EA	2	1083	
Controls	AA	2	3755	6706
	EA	0	2951	
Total	AA	6	5347	9381
	EA	2	4034	

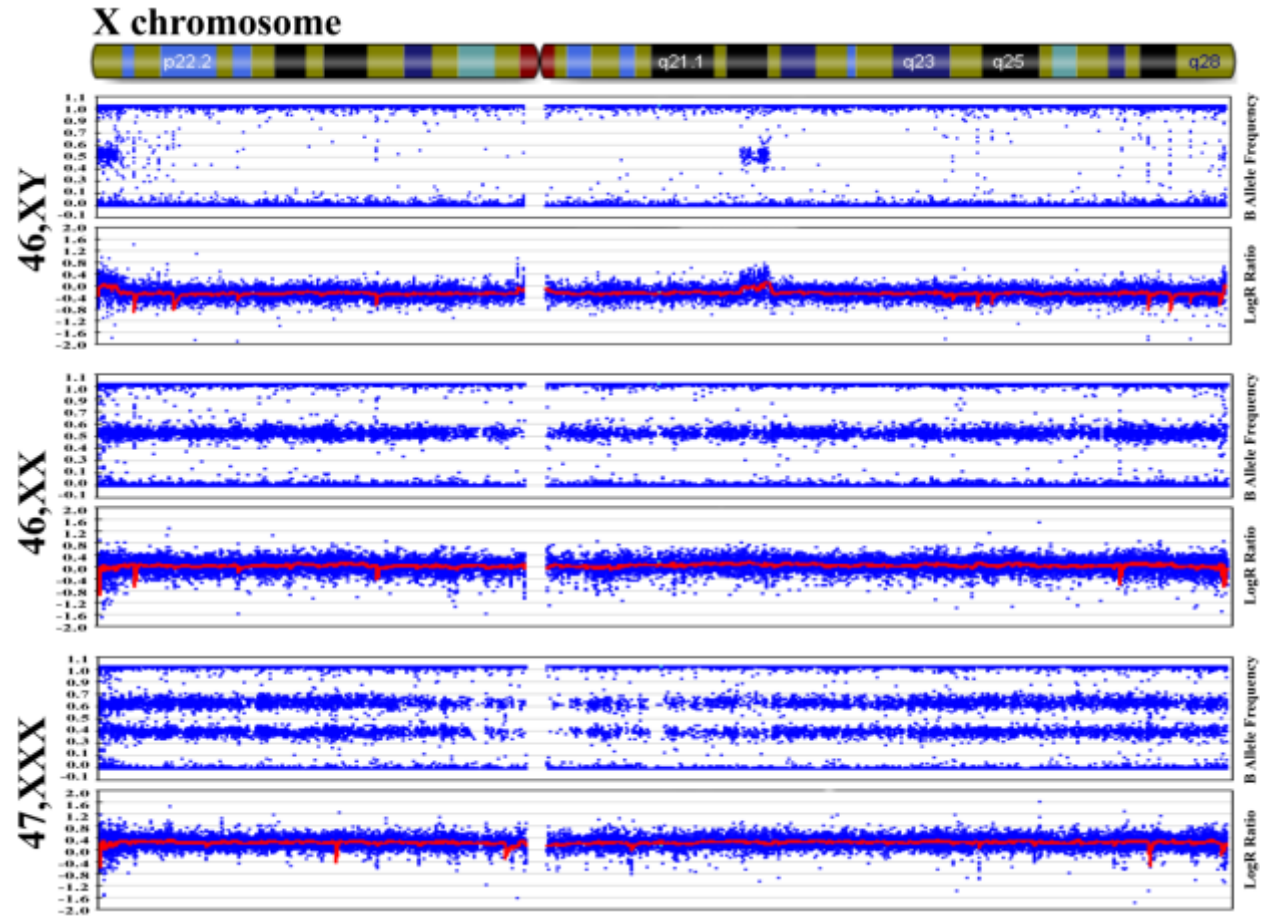
**Table 3. Systemic lupus erythematosus (SLE) and primary Sjögren's syndrome (SS)**  
**relative risk with X chromosome numbers.**

<b>X number</b>	<b>Karyotype</b>	<b>SLE Relative Risk</b>	<b>SS Relative Risk</b>
1	45X; 46XY	1	1
2	46XX; 47XXY*	~10	~14
3	47XXX	~25	~41

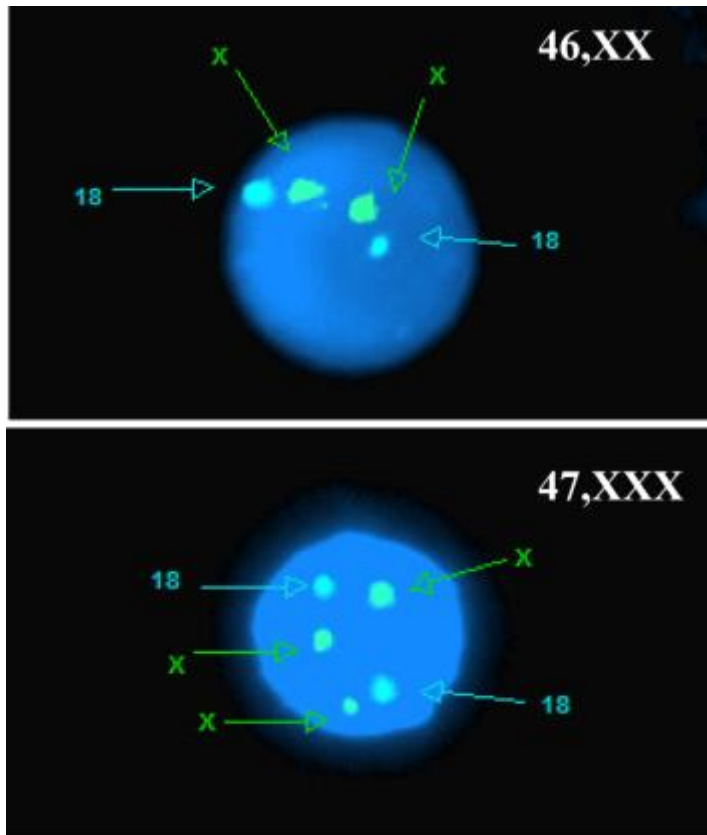
---

\*47,XXY data with Sjögren's syndrome is not available here.

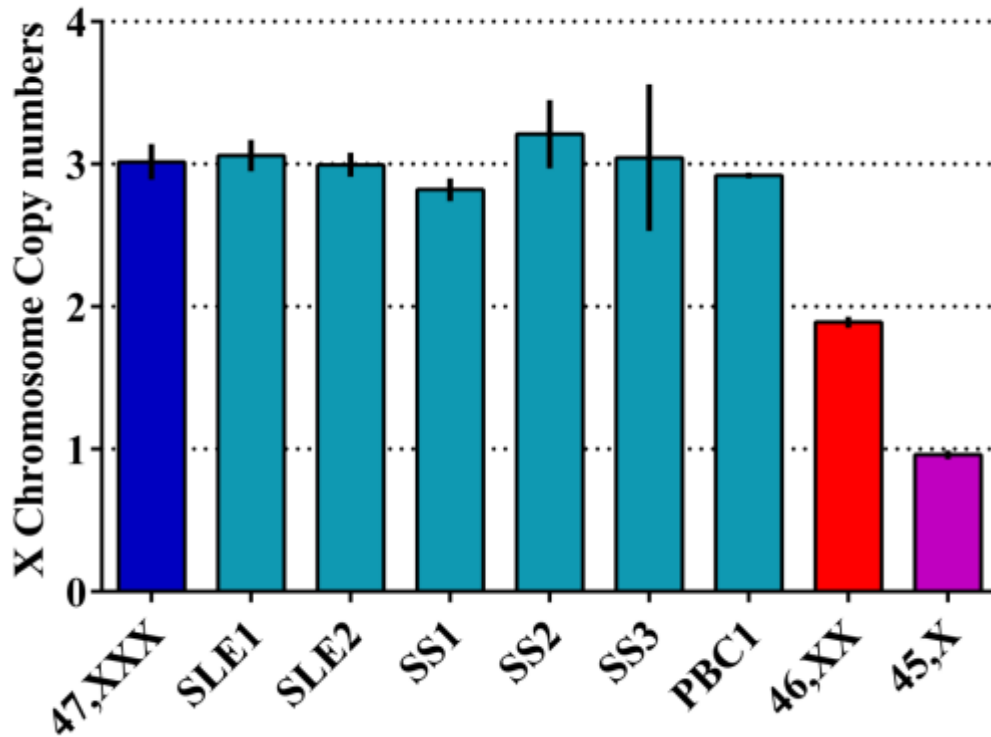
## Figures



**Fig. 1. 47,XXX identification.** B allele frequency and logR ratio plot of 46,XY from a representative normal man (first panel, for comparison purposes only as men were not included in the present study), 46,XX from a normal woman (second panel) and 47,XXX from a trisomy X syndrome patient (third panel).



**Fig. 2. 47,XXX validation by fluorescence *in situ* hybridization assay (FISH).** Validation of 46,XX from a normal female (upper) and 47,XXX from a trisomy X syndrome patient (bottom). The images are showing a single representative nucleus. 200 nuclei were counted.



**Fig. 3. 47,XXX validation by PCR.** DNA amplification validation of 47,XXX for two systemic lupus erythematosus patients (SLE1, SLE2), two primary Sjögren's syndrome patients (SS1, SS2) and one primary biliary cirrhosis patient (PBC1). As calibrators, we used known 47,XXX (shown in blue), 46,XX (shown in red) and 45X (shown in purple) samples. Calibrators were all validated by FISH and used to determine copy number in the experimental samples.

**Acknowledgments:**

We thank for AECG for SS data. We thank for Dr. Siminovitch's group for kindly providing us PBC and part of the RA data. We thank for GENAR collaboration for providing us part of the RA data. We thank for Dr. Montgomery's group for providing us the sarcoidosis data. We thank the Cincinnati control cohort group for allowing us to use the data. We thanks Dr. Bottinger for providing us the Mt. Sinai African American control dataset. Also, this research would not have been possible without the help and cooperation of the patients and others who participated in this study. We thank the cohort personnel, physicians, and genotyping core for their help with acquisition, referral, and processing of several thousand samples.

**Author contributions:**

JB Harley and RH Scofield designed this study conception.

K Liu, BT Kurien, SL Zimmerman, KM Kaufman, DH Taft, JB Harley and RH Scofield participated in data analysis.

K Liu, BT Kurien, KM Kaufman, LC Kottyan, S. Lazaro, CA Weaver, JA Ice, AJ Adler, J Chodosh, L Radfar, A Rasmussen, DU Stone, KS Hefner, GD Houston, DM Lewis, S Li, KA Koelsch, A Igoe, M Dave, J Kumar, A D'souza, JS Maier-Moore, VM Harris, R Gopalakrishnan, R Jonsson, JA Lessard, X Lu, JE Gottenberg, JM Anaya, DS Cunninghame-Graham, AJW Huang, MT Brennan, P Hughes, GG. Illei, C Miceli-Richard, EC Keystone, VP Bykerk, G Hirschfield, G Xie, WF Ng, G Nordmark, P Eriksson, R Omdal, NL Rhodus, M Rischmueller, M Rohrer, BM Sega, TJ Vyse, M Wahren-Herlenius, T Witte, X Mariette, B Pons-Estel, ME Alarcon-Riquelme, JM Guthridge, JA James, CJ Lessard, JA Kelly, SD Thompson, PM Gaffney, CG Montgomery, KA Siminovitch, JC Edberg, RP Kimberly, GS Alarcón, CL Langefeld, GS

Gilkeson, DL Kamen, BP Tsao, WJ McCune, JE Salmon, JT Merrill, MH Weisman, DJ Wallace, TO Utset, EP Bottinger, CI Amos, KL Sivils, JB Harley and RH Scofield participate in data acquisition.

K Liu, DH Taft, JB Harley and RH Scofield worked on manuscript preparation.

**Supplementary Table 1a. Recruitment information of patient datasets.** Pre- and post-quality control (QC) numbers, genotyping platforms, recruitment strategies, criterias, locations, time and references are listed for SLE, SS, PBC, RA and sarcoidosis patient datasets. Age information is unavailable.

Datasets	Pre QC	Post QC	Genotyping platform	Recruitment strategy	Recruitment Criteria	Recruitment locations	Recruitment time	Reference
SLE	3449	2826	Omni 1, and immunochip	Consent in clinics	American College of Rheumatology (ACR) classification criteria for systemic lupus erythematosus	Cedars-Sinai Medical Center, California; Hospital for special surgery, New York; Oklahoma Medical Research Foundation, Oklahoma; University of Alabama at Birmingham, Alabama; Univeristy of Chicago, Illinois; University of Michigan, Michigan; University of Minnesota, Minnesota; United states; 27 subjects unknown	976 subjects: before 2010; 1821 subject: 1992-2010; 29 subjects unknown	Not published
SS	1238	1033	Immunochip	Consent in clinics	American-European Consensus Group (AECG) classification criteria for primary Sjögren's syndrome	Newcastle University, United Kingdom; National Institue of Health, USA; Uppsala University Hospital, Sweden; Karolinska Institute, Sweden; The Queen Elizabeth Hospital and University of Adelaide, Australia; Stavanger University Hospital, Norway; University of Bergen, Norway; Université Paris-Sud, France	N/A	(31)
PBC	1564	1118	HumanHap 370 Beadchip; Immunochip	Consent in clinics	American Association for the Study of Liver Diseases Diagnostic Criteria for PBC	Canada, United States and Europe	2005-2012	(47, 48)
RA_1	453	449	HumanHap 370 Beadchip	Consent in clinics	1987 American College of Rheumatology criteria for the classification of RA	Toronto, Canada	2004-2008	(30)
RA_2	1290	1261	Immunochip	Consent in clinics	1987 American College of Rheumatology criteria for the classification of RA	Argentina, Mexico, Peru and Chile	2008-2010	(49)
Sarcoidosis	956	939	Omni 1	Consent in clinics, recruit from health fairs, advertise through media, hospital and church	Tissue confirmation of granuloma; or had chest radiographic evidence of bilateral symmetrical hilar adenopathy and either a history of erythema nodosum or at least two years observation during which time no other disease was found to explain radiographic abnormalities	Beth Israel Deaconess Medical Center, Massachusetts; Georgetown University Medical Center, Washington D.C.; Case Western Reserve University and University of Cincinnati Medical Center, Ohio; Henry Ford Health system, Michigan; Johns Hopkins University School of Medicine, Maryland; Medical University of South Carolina, South Carolina; Mount Sinai Medical Center, Florida; National Jewish Medical and Research Center, Colorado; University of Iowa College of Medicine, Iowa; University of Pennsylvania and Allegheny University of the Health Sciences, Pennsylvania; United States. 429 subjects unknown.	167 subject: 1996-1999; 429 subjects: late 90's to early 2000's; 343 subjects: unknown	(23-26)



**Supplementary Table 1b. Recruitment information of control datasets.** Pre- and post-QC numbers, genotyping platforms, recruitment strategies, criterias, locations, time and references are listed for control datasets. Age information is unavailable.

Datasets	Pre QC	Post QC	Genotyping platform	Recruitment strategy	Recruitment criteria	Recruitment locations	Recruitment time	Reference	
Non-affected SLE family controls	2408	2090	Immunochip	Consent in clinics	First degree lupus family members (mothers and sisters) who have no SLE	United States	1990-2010	(27)	
				Community-based recruitment strategies: including visits to neighborhood schools and day-care centers	Healthy individuals without known major health conditions recruited from the geographic area served by Cincinnati Children's Hospital Medical Center (CCHMC)	Cincinnati, OH, United States	Sept 2006-Oct 2010	(28) (29)	
Healthy individuals, with no auto-inflammatory diseases	RDRCC cohorts	792	747	Immunochip	Consent in community clinics, recruit in health fair and employer sponsored recruiting events	Healthy individuals recruited who were screened negative for autoimmune rheumatic disease by a validated questionnaire(50)	Oklahoma, United States	2009-2010	Not published
	Canada cohort	452	440	HumanHap 370 Beadchip, Immunochip	Individuals of European origin from the Toronto area who had no history of rheumatoid arthritis or other inflammatory disease	Toronto, Canada	2005-2011	(30), data for 112 subjects not published	
	SS controls	465	352	Immunochip	N/A	Population controls	Uppsala University Hospital, Sweden; Karolinska Institute, Sweden; The Queen Elizabeth Hospital and University of Adelaide, Australia; Oklahoma Medical Research Foundation, USA; Stavanger University Hospital, Norway; University of Bergen, Norway; University of Minnesota, USA	N/A	(31)
Population Controls	AA controls	2379	2328	OmniExpress	Consent in clinics	African American patients receiving care at Mount Sinai without selection for any specific condition requirements	Mt Sinai Medical School hospitals	2007-2012	Not published
				Consent in clinics, recruit in health fairs, advertise through media, hospital and church	Non-sarcoidosis	Beth Israel Deaconess Medical Center, Massachusetts; Georgetown University Medical Center, Washington D.C.; Case Western Reserve University and University of Cincinnati Medical Center, Ohio; Henry Ford Health system, Michigan; Johns Hopkins University School of Medicine, Maryland; Medical University of South Carolina, South Carolina; Mount Sinai Medical Center, Florida; National Jewish Medical and Research Center, Colorado; University of Iowa College of Medicine, Iowa; University of Pennsylvania and Allegheny University of the Health Sciences, Pennsylvania; United States. 106 subjects unknown.	186 subject: 1996-1999; 105 subjects: late 90's to early 2000's; 329 subjects: unknown	(23-26)	
Sarcoidosis controls	639	620	Omni 1						

**Supplementary Table2. Adjusted and unadjusted OR of 47,XXX in SLE.**

<b>Logistic regression modeling</b>	<b>OR</b>	<b>p-value</b>
Model 1 (47,XXX only)	7.53	0.01
Model 2 (47,XXX and ancestry)	7.36	0.01

**Supplementary Table 3. 47,XXX enrichment in SLE (compared to non-affected SLE family members).** 47,XXX in systemic lupus erythematosus (SLE) and non-affected SLE family controls. \*Modified p-value, OR with 95% CI were calculated by fisher exact test through adding 0.5 into each cell to solve the issue of zero 47,XXX in the controls.

<b>Disease phenotype</b>	<b>Sample Size</b>	<b>47, XXX</b>	<b>p-value</b>	<b>OR(95% CI)</b>
Systemic lupus erythematosus	2826	7	0.02	11.12 (1.26-inf)
Non-affected SLE family controls	2090	0	-	-

**Supplementary Table 4. 47,XXX enrichment in SS (compared to healthy non-auto-inflammatory controls).** 47,XXX in Sjögren's syndrome (SS) and healthy non-auto-inflammatory controls. \*Modified p-value, OR with 95% CI were calculated by fisher exact test through adding 0.5 into each cell to solve the issue of zero 47,XXX in the controls.

<b>Disease phenotype</b>	<b>Sample Size</b>	<b>47, XXX</b>	<b>p-value</b>	<b>OR(95% CI)</b>
Sjögren's syndrome	1033	3	0.02	11.44(1.08-inf)
Healthy non-auto-inflammatory controls	1684	0	-	-

**Supplementary Table 5a. 47,XXX enrichment in SLE and SS (compared to controls without non-sarcoidosis control cohort).** 47,XXX in systemic lupus erythematosus (SLE), primary Sjögren's syndrome (SS) and controls without non-sarcoidosis controls. Each disease state is compared to controls without non-sarcoidosis controls. P-value, OR with 95% CI were calculated by Fisher exact test.

<b>Disease phenotype</b>	<b>Sample Size</b>	<b>47, XXX</b>	<b>p-value</b>	<b>OR(95% CI)</b>
Systemic lupus erythematosus	2826	7	0.001	16.02 (2.06-719.67)
Sjögren's syndrome	1033	3	0.009	18.78 (1.51-980.69)
Controls (No non-sarcoidosis)	6454	1	-	-

**Supplementary Table 5b. 47,XXX enrichment in SLE and SS (compared to controls**

**without population control cohort).** 47,XXX in systemic lupus erythematosus (SLE), primary Sjögren's syndrome (SS) and controls without population controls. Each disease state is compared to controls without population controls. P-value, OR with 95% CI were calculated by Fisher exact test.

<b>Disease phenotype</b>	<b>Sample Size</b>	<b>47, XXX</b>	<b>p-value</b>	<b>OR(95% CI)</b>
Systemic lupus erythematosus	2826	7	0.007	10.90 (1.40-490.58)
Sjögren's syndrome	1033	3	0.02	12.79 (1.03-669.11)
Controls (No population controls)	4394	1	-	-

**Supplementary Table 5c. 47,XXX enrichment in SLE and SS (compared to controls**

**without non-sarcoidosis and population control cohorts).** 47,XXX in systemic lupus erythematosus (SLE), primary Sjögren's syndrome (SS) and controls without both non-sarcoidosis controls and population controls. Each disease state is compared to controls without non-sarcoidosis and population controls. \*Modified p-value, OR with 95% CI were calculated by fisher exact test through adding 0.5 into each cell to solve the issue of zero 47,XXX in the controls.

<b>Disease phenotype</b>	<b>Sample Size</b>	<b>47, XXX</b>	<b>p-value*</b>	<b>Modified OR(95% CI)*</b>
Systemic lupus erythematosus	2826	7	0.001	20.08 (2.28-inf)
Sjögren's syndrome	1033	3	0.002	25.64 (2.41-inf)
Controls (No non-sarcoidosis and population controls)	3774	0	-	-

**Supplementary Table 6. 47,XXX incidence (compared to unselected newborn infants cohort).** 47,XXX in systemic lupus erythematosus (SLE), primary Sjögren's syndrome (SS), primary biliary cirrhosis (PBC), rheumatoid arthritis (RA), sarcoidosis and unselected newborn infants. Each disease state is compared to the unselected newborn infants group. P-value, OR with 95% CI were calculated by Fisher exact test.

<b>Disease phenotype</b>	<b>Sample Size</b>	<b>47, XXX</b>	<b>OR(95% CI)</b>	<b>p-value</b>
Systemic lupus erythematosus	2826	7	2.58 (0.92-6.36)	0.036
Sjögren's syndrome	1033	3	3.02 (0.57-10.22)	0.093
Primary biliary cirrhosis	1118	1	0.93 (0.02-5.82)	1
Rheumatoid arthritis	1710	1	0.61 (0.01-3.80)	1
Unselected newborn infants(34)	20790	20	-	-